The Future Role of Machine Learning in Clinical Transplantation

Review

The Future Role of Machine Learning in Clinical Transplantation

Katie L. Connor, MRCS,[1,2,3] Eoin D. O'Sullivan, MBBS,[3] Lorna P. Marson, MD,[2,3] Stephen J. Wigmore, MD,[2,3] and Ewen M. Harrison, PhD[4]

Abstract: The use of artificial intelligence and machine learning (ML) has revolutionized our daily lives and will soon be instrumental in healthcare delivery. The rise of ML is due to multiple factors: increasing access to massive datasets, exponential increases in processing power, and key algorithmic developments which allow ML models to tackle increasingly challenging questions. Progressively more transplantation research is exploring the potential utility of ML models throughout the patient journey, although this has not yet widely transitioned into the clinical domain. In this review, we explore common approaches used in ML in solid organ clinical transplantation and consider opportunities for ML to help clinicians and patients. We discuss ways in which ML can aid leverage of large complex datasets, generate cutting-edge prediction models, perform clinical image analysis, discover novel markers in molecular data, and fuse datasets to generate novel insights in modern transplantation practice. We focus on key areas in transplantation where ML is driving progress, explore the future potential roles of ML, and discuss the challenges and limitations of these powerful tools.

(*Transplantation* 2020;00: 00–00).

INTRODUCTION

From the beginning of solid organ transplantation, there has been a fastidious approach to the collection of data. This highly organized approach has allowed important observational studies to be performed and has provided rich data for traditional research methods. However, we have now entered an era where these increasingly large and integrated datasets—including electronic health records, clinical images, and multiomics datasets—are amenable to more sophisticated machine learning (ML) approaches.

We use ML every day, often without realizing it. We might log onto a phone using facial recognition, issue requests to a virtual personal assistant using our voice (eg, Siri or Alexa), or compose sentences in an email with predictive natural language processing. ML tools have also become central to modern biomedical research, due to better access to large datasets, the exponential increase in processing power, and key algorithmic developments allowing ML models to tackle increasingly challenging data. In healthcare, ML models can already harness electronic healthcare records to predict the future risk of acute kidney injury,[1] diagnose retinopathy,[2] and derive novel features from existing data, such as the detection of patients with atrial fibrillation from an ECG acquired during sinus rhythm[3] and the deduction of coronary arterial blood flow from cardiac CT.[4]

In this review, we explore how ML can now help clinicians and patients to develop cutting-edge prediction models, perform clinical image analysis, analyze high-throughput genomics and analyze complex fused datasets, with the potential to generate novel insights relevant to modern transplantation practice (Figure 1). We will focus on some key clinical themes in transplantation where ML is driving progress, explore the future clinical potential of ML and discuss the challenges and limitations of these powerful tools.

ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, AND DEEP LEARNING

Artificial intelligence (AI) can been loosely be described as the ability to automate and enhance intellectual tasks normally performed by humans, in a way that we

Received 11 March 2020. Revision received 30 June 2020.

Accepted 21 July 2020.

[1] *Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, United Kingdom.*

[2] *Edinburgh Transplant Unit, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom.*

[3] *Centre for Inflammation Research, University of Edinburgh, Edinburgh, United Kingdom.*

[4] *Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom.*

The authors declare no conflicts of interest.

MRC/KRUK Clinical Research Training Fellowship - Ref MR/S001743/1 to K.L.C. KRUK Clinical Research Training Fellowship - TF_006_20161125 to E.D.O.S.

K.L.C. is involved in conceptualization, performance of the research, writing and editing the manuscript. E.D.O.S. is involved in performance of the research, writing and editing the manuscript. L.P.M. is involved in supervision, research design, editing the manuscript. S.J.W. is involved in supervision, conceptualization, research design, editing the manuscript. E.M.H. is involved in supervision, conceptualization, performance of the research, writing and editing the manuscript.

Correspondence: Katie L. Connor, MRCS, Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, EH16 4TJ, United Kingdom. (katie.connor@ed.ac.uk).

ISSN: 0041-1337/20/00XXX-00

DOI: 10.1097/TP.0000000000003424

<zdoi;10.1097/TP.0000000000003424>

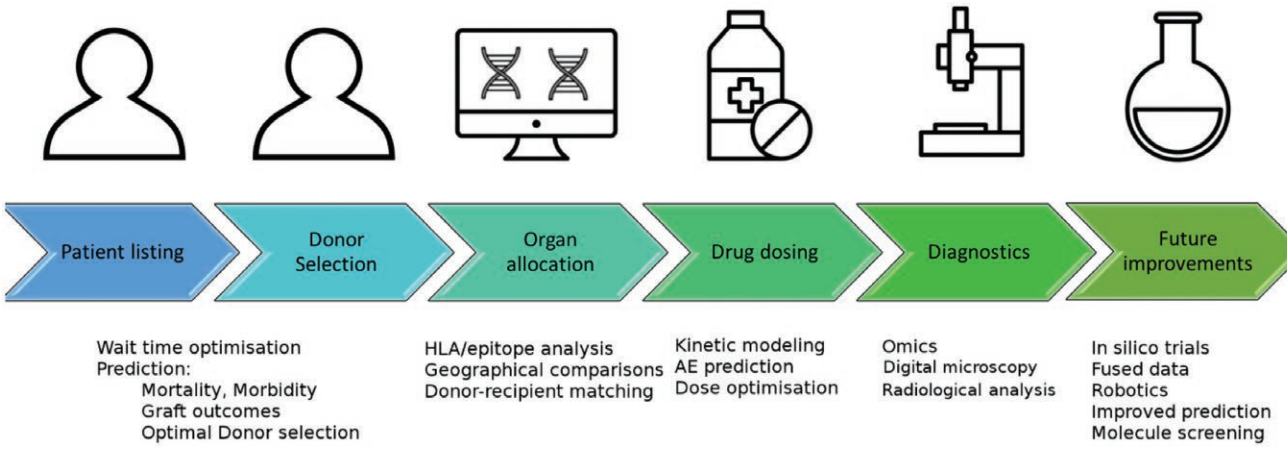


FIGURE 1. Applications of machine learning approaches to clinical transplantation. ML approaches have been tested at multiple stages in the journey of a transplant recipient with the ability to analyze large and diverse datasets. In future, we may see increasing use of ML for these roles and in transplantation research. AE, adverse events.

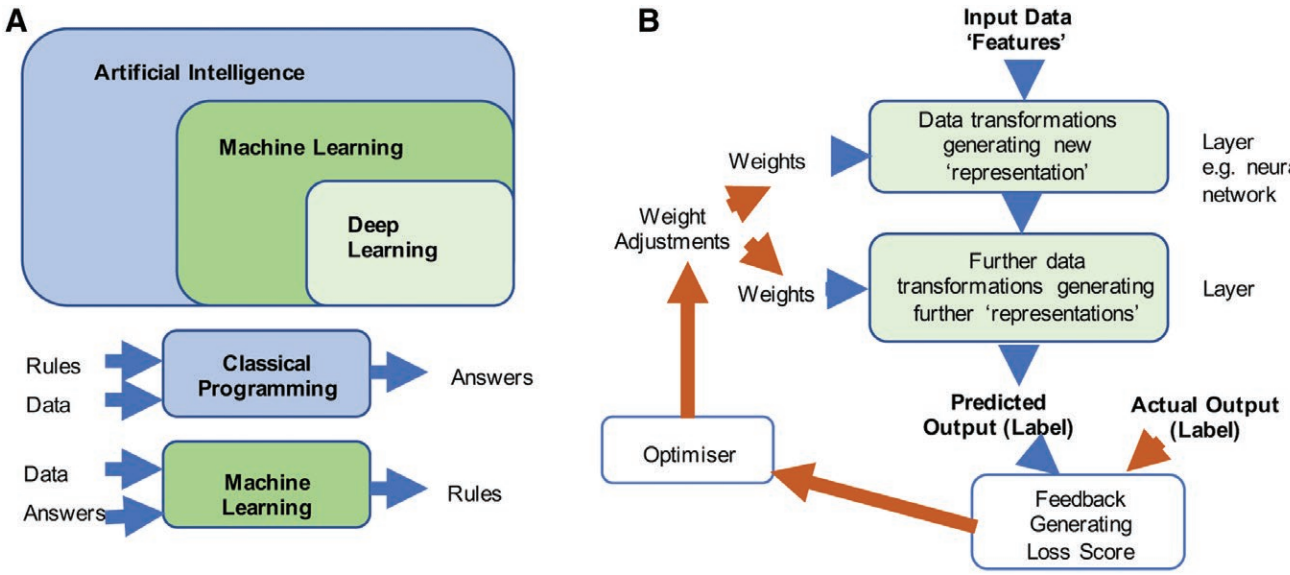


FIGURE 2. The core concepts in machine learning and deep learning. A, Machine learning (ML) and deep learning (DL) are both forms of artificial intelligence. The key difference is that in ML and DL, the models are capable of generating their own "rules" or data transformations constituting the training or "learning" component of the model which can later be applied to datasets where the expected answer is not known. B, Basic structure of the components of a DL model and the iterative structure of training. During the training, the user provides the input data which will comprise a number of variables (features). These data will undergo a series of transformations and the model will generate a prediction (eg, graft failure). The model performance is then assessed by comparing the predicted outcome with the actual outcome (Loss score). This is then fed back into the model to optimize the data transformation steps (eg, by modifying the weighting given to a feature). This process is then repeated until the predicted outcome matches the actual outcome as closely as possible. This ML algorithm can then be applied to "unlabeled" data where the outcome is not known to generate a prediction. Figures adapted from Deep Learning with R.[5]

consider "smart." ML is one important facet of AI, with which a machine can "learn" from data without being explicitly programmed to do so. At its most fundamental, ML requires humans to provide both the input data (or features) and to define an expected output. Thereafter, the ML procedure generates a set of rules which allow the input to lead to the desired output (Figure 2). Often, this process is performed iteratively until the model can achieve a predicted output which matches the provided actual output or "label" with a desired level of speed and accuracy; this constitutes the "learning" or training component of the model. Once trained, the ML system should be able to generate similarly accurate outputs from new unlabeled data based on these learned "rules." This learning can be thought of as a series of data transformations, performed to generate more contextually meaningful representations with the goal of matching input to output.[5] By generating algorithmic "rules" to interpret data, these systems can adapt to analyze complex datasets—an ability of great potential for healthcare data analysis.

To help understand when ML might be applicable to solving clinical questions, it can be helpful to consider approaches under 3 broad categories: supervised learning, unsupervised learning, and deep learning (DL). The decision on the most appropriate approach depends on the research question (Table 1) and, in practice, many workflows will have elements of multiple categories.

TABLE 1.

Examples of machine learning approaches and their application

ML approach	Suitable questions	Example algorithms
Supervised learning	Will the clinical features in my data predict kidney transplant eGFR at 1-year post-transplantation? (a regression problem)	Regression: Logistic regression, ridge regression, and regressions with LASSO (least absolute shrinkage and selection operator), support vector regression
	Can I use laboratory data in an ITU cohort to identify patients who will require liver transplant within 1 mo? (a classification problem)	
	Does this kidney transplant biopsy demonstrate rejection and stage (eg, according to the Banff Classification) (a multi-class classification problem)	Classification and clustering: Naive-Bayes classifiers, decision trees, random forests, and stochastic gradient boosting machines
Unsupervised learning	Given a set of patient genomic data, cluster into similar genotypes (clustering)	Graph-based clustering
	Given a set of single-cell transcriptomes, cluster into similar subtypes (clustering)	tSNE UMAP K-Means clustering
Deep Learning	Given a training dataset of CT scans classified as malignancy/no malignancy, learn to classify new CT scans	Artificial Neural Networks (ANNs)
	Given a training transcriptomic dataset of where an immune cell of interest is classified, identify these cells in new data	Convolutional Neural Networks (CNN)
	Given a large integrated dataset of longitudinal clinical, biochemical and genomic data and rejection outcomes, predict which patients will suffer organ rejection in new test data	Recurrent Neural Networks (RNN)

tSNE, T-distributed stochastic neighbor embedding; UMAP, Uniform Manifold Approximation and Projection.

In supervised learning, the user can provide an initial dataset, where the "outcomes" are classified (labeled), and the algorithm can learn to make predictions using the input data (features). Here, the term supervised refers to the fact that we provide labels for the desired outcome (eg, graft survival), guiding the learning of the model. Once trained, the model should be able to predict outcomes based on the input features alone (unlabeled data). By contrast, in unsupervised learning, the model is trained on unlabeled data. Here the output is achieved by "clustering" data and searching for similarities or heterogeneity. For example, when analyzing molecular features of a biopsy, the model will search for and extract features from the data allowing clustering into clinical subtypes. This has the benefit of not requiring a researcher to pre-label features which might be either unfeasible, prohibitively labor-intensive, or introduce excessive bias. An unsupervised approach might also support the discovery of hidden or novel features or structure within data.

"DL" is a new branch of ML that has emerged as an important tool for analyzing complex data such as images, speech, and language. In DL, successive sequences of data transformations, termed layers, are employed to increase a model's accuracy and generate meaningful representations of the data (Figure 2). It is these "layers" that represent the "deep" in DL. DL models are based on artificial neural networks (ANNs). Neural networks are networks of interconnected nodes which are modeled loosely on the neurons in the human brain. Each "node" can be assigned a value based on the features of the dataset, which can be tuned and weight-adjusted (typically by the network itself as it trains itself) before being passed to the next node (Figure 2). This may occur thousands of times or more as the models retain those elements that increase accuracy and adjust others, thus slowly converging on an optimal solution. The number of these nodes, their overall structure, and the specifics of how they connect to each other is known as the network architecture. There are many architectures under the umbrella term of ANN including convolutional neural networks (CNNs), recurrent neural networks, and general adversarial networks. Readers may wish to experiment with the TensorFlow playground, an opensource library of tools for ML, which allows interactive visualizations of neural networks and how they operate (https://playground.tensorflow.org).

APPLICATIONS OF ML IN TRANSPLANTATION

Clinical Prediction and Decision Support

ML can be used as a decision support tool by transplant clinicians at many points in the patient journey (Figure 1). Whilst regression models remain a standard approach when trying to predict outcomes in patients, increasingly complex supervised ML methods and DL approaches are being trialed with some success.[6,7] There are several perceived advantages of these more complex ML models. First, their ability to handle billions of raw input variables is an advantage when analyzing complex heterogenous datasets such as electronic health records. Applied to registry data, ML models can theoretically better handle nonlinear predictor-outcome relationships and multidimensional interactions between predictors—allowing them to find uncommon patterns and novel predictors of outcomes within the data.

Listing for Transplantation

Predicting waiting list mortality and morbidity is important in transplant decision-making in both listing for transplantation, organ allocation algorithms and determining whether to accept a donor organ for a specified recipient. ML has helped researchers in cardiac transplantation, where random forest analyses have been employed to identify novel features that confer additive risk through complex interactions, not previously seen using Cox proportional hazards models.[8] However, many variables relevant to waiting list mortality were not captured in the registry data and hence comparison against the gold standard was not possible.[8]

In liver transplantation, an ML classification tree was deployed to create an optimized prediction of mortality (http://www.opom.online) to predict a patient's 3-month waiting list mortality.[6] The model outperformed Model for End-Stage Liver Disease scores, and allocation based on optimized prediction of mortality would theoretically reduce mortality by 418 deaths annually in a simulated liver allocation model analysis. This apparent improvement justifies exploring the potential of such approaches in other organs and considering their role in clinical practice and waiting list management.

Organ Allocation

Allocating organs is a complex process with multiple stakeholders, involving intensive computational elements, while incorporating ethical and political considerations. It is understandably impossible to test multiple allocation schemes simultaneously in real life; however, ML enables a variety of approaches to be modeled in silico. Multiperiod linear optimization modeling has allowed variations of kidney sharing practice and outcomes to be compared, which could facilitate novel approaches and computational changes to be tested before implementation, informing allocation strategies in the future.[9] Simulated matching schemes allow for optimization for the desired patient and graft variables and even predict regional deficits. Neural networks have been successfully used to test a variety of donor-recipient matching approaches for liver transplantation.[10,11] ML allows us to hold up a mirror to our own decisions in this area, where human variability in decision making as to whether to accept a lung for transplantation has been captured using fuzzy decision tree models (82.1% accuracy).[12] Finally, in silico modeling has allowed nonpermissive HLA mismatching to be predicted in hematopoietic cell transplantation based on an amino acid polymorphism on T-cell alloreactivity.[13] Similar ML approaches may yet inform future solid organ allocation schemes with the potential ability to rapidly integrate complex emerging elements of epitope matching, such as rapid protein-protein interaction modeling.

Prediction of Patient and Graft Survival

The majority of published clinical transplantation papers employing ML concern the prediction of either patient or graft survival following transplant, as summarized in Table 2. Typically, these studies analyze large registry datasets using a variety of ML methods, most commonly tree-based methods (decision/classification trees, random forests), Bayesian networks, and ANNs, with performance usually compared to traditional regression models or known clinical prediction scores.[14]

The accurate prediction of future graft function remains particularly challenging in cardiac transplantation, and this contributes to high organ discard rates.[15] Over the last decade, increasingly sophisticated ML models have been employed to better predict cardiac graft function.[7,8,16-18] Initially, studies primarily employed ML models to identify novel predictive variables, which were then analyzed using logistic regression approaches.[18] The International Heart Transplantation Survival Algorithm (http://www.ihtsa.med.lu.se) employs ANN, decision-trees, and regression, outperforming standard classification scoring (eg, Donor Risk Index and index for mortality prediction after cardiac transplantation) with an area under the receiver operating curve (AUROC) 0.650 for the prediction of 1-year survival.[7] The IHSTA underwent external validation with Nordic data during development and has been subsequently tested on new United Network for Organ Sharing (UNOS) data with sustained performance.[17] The Tree of Predictors model, provides risk stratification tailored to specific patient clusters and achieves an AUROC of 0.660 for 3-month survival on UNOS data.[8] The next logical step would be for the direct comparison of these scoring algorithms on previously unseen data.

Despite this promising work there remains significant debate as to whether the marginal gains in predictive ability offered by these models are worth the limitations of their inherent complexity. Miller et al[16] compared a number of regression models with multiple ML methodologies, including neural networks, naive-Bayes, tree-augmented naive-Bayes, support vector machines (SVMs), random forest, and stochastic gradient boosting, and found no significant advantage over regression analyses in predicting 1-year survival after cardiac transplantation from the UNOS registry data. The authors concluded that where features (eg, registry data) are well defined, there may be minimal differences between complex classifiers (eg, deep neural networks, random forests) as compared to similar classifiers (eg, logistic regression).[16]

In liver transplantation, ML models have been employed to predict graft failure and patient survival in both adult and pediatric recipients (Table 2).[19-21] In adult patients undergoing liver transplantation, Molinari et al[19] were able to very accurately predict patients with a high chance (≥10%) of mortality based on preoperative features using the combination of classification trees and ANNs (AUROC = 0.952). Whilst this study could be considered in decision support for listing, a limitation is that it only includes patients who underwent surgery, rather than encompassing all patients who were potentially eligible for OLT. Other studies have been shown to either complement or outperform standard scoring algorithms such as Model for End-Stage Liver Disease.[20,21] Following liver transplantation, Lee et al[22] were able to accurately predict which patients developed acute kidney injury using SVM and random forest analyses, potentially providing a window for targeted intervention in high-risk patients.

Whilst in renal transplantation, a large number of studies have focused on the prediction of outcomes including delayed graft function, graft rejection, and graft survival (Table 2). The ability of such models to correctly identify a graft that will survive is typically quite high (AUROC > 0.8); however there is notable variation in performance even using the same models, suggesting that the training data is of vital importance in such cases.[23-27] It remains unclear whether these ML models can outperform those based on multivariate regression methods, such as the iBox risk prediction score[28] which is currently being considered by the Food and Drug Administration (FDA) for approval for use as a surrogate endpoint in clinical trials.

Following transplantation, the likelihood of future events or graft function is based on the trajectory/events to that point. For example, if an episode of acute rejection (AR) occurs in the first year after transplant, the future risk of renal transplant failure is higher.[29] The current transplant predictive scores are heavily biased to preoperative

TABLE 2.

Summary of studies employing machine learning in the prediction of either patient survival or graft outcomes following transplantation

Author (year)	Prediction	Data	ML model used	Result
Kidney				
Yoo et al (2019)[93]	Graft survival	Multicentre Study, Korea, 1997–2012, (n = 3117)	Multiple Models Inc.: Tree-based models	Survival decision tree model performed best (C-Index 0.80) for 10-y survival
Mark et al (2019)[94]	Graft survival	UNOS, 2002–2011 (n = 163 199)	Tree-based model	5-y C-Index 0.724 Outperformed the Estimated Post Transplant Survival (EPTS) score
Shaikhina et al (2019)[95]	Graft rejection	Patient records, United Kingdom, 2003–2012 (n = 80), HLA incompatible patients	Tree-based model	85% accuracy. Identified factors associated with rejection
Topuz et al (2018)[96]	Graft survival	UNOS 2004–2015 (n = 31 207)	Feature Selection Inc: SVMs, ANN, tree-based models Analysis: BBN	Highest accuracy achieved using the BBN (accuracy 0.684) on fused data mining model
Tapak et al (2017)[97]	Graft Rejection	Patient records, Iran, 1994–2011 (n = 378)	Artificial Neural Network	AUROC ANN: 0.88, LR: 0.75
Nematollahi et al (2017)[98]	Graft survival	Patient records, Iran, 2002–2007 (n = 717)	Multiple compared: ANN, SVM	AUROC: SVM = 0.86, ANN = 76.9, LR = 77.4
Shahmoradi et al (2016)[99]	Graft survival	Patient data, Iran, 2007–2013 (n = 513)	Multiple compared: ANN, tree-based models	Overall accuracy: ANN 83.7%, tree-based models 83.28%–87.21%
Esteban et al (2016)[30]	Death, graft loss, rejection	Patient data, Germany, 2005 (n = 2061)	Artificial Neural Network	Recurrent Neural Network performed best: AUROC 0.77–0.89
Fouad et al (2015)[100]	Graft survival	Patient records, Egypt, 1976–2007, live donor recipients	Regression, tree-based model, rule-based classifiers	Correlation coefficient 0.0.87, 0.737, and 0.733 for ruled based, TBM, and regression, respectively
Decruyenaere et al (2015)[101]	Delayed graft function	Patient records, 2005–2011 (n = 497), Belgium	Multiple models Inc: SVM, tree-based models	Linear SVM has the highest discriminative capacity (AUROC of 84.3%) outperforming regression
Brown et al (2012)[102]	Graft survival	US Renal Data System, 2000-2001 (n = 5144)	Bayesian belief network	1-y AUROC = 0.59, 3-y AUROC = 0.63
Lasserre et al (2012)[103]	1-y eGFR	Eurotransplant (n = 707), 1998–2008	SVM	Pearson correlation coeff. between predicted and real eGFR of 0.48
Li et al (2010)[104]	Graft survival	Regional Patients/UNOS, 1987–2009 (n = 1228)	Bayesian belief network	Model could predict graft status but not duration of graft survival
Lofaro et al (2010)[105]	Chronic allograft nephropathy	Italy, patient records (n = 80)	Tree based model	5-y AUROC = 0.847
Hummel et al (2010)[27]	Acute rejection, nephrotoxicity	Patient data, Brasil (n = 145)	Artificial Neural Network	ARej: best AUROC = 0.78 Nephrotoxicity: best AUROC = 0.66
Greco et al (2010)[106]	Graft survival at 5 y	Patient data, Italy (n = 194)	Tree-based model	88% sensitive, 73% specific
Akl et al (2008)[26]	Graft survival	Patient-level, Egypt	Artificial Neural Network	ANN predictive accuracy was 88% vs regression approach (72%)
Santori et al (2007)[25]	Delayed graft function	Patient records, Italy, pediatric recipients (n = 148)	Artificial Neural Network	ANN: sensitivity 0.875 and specificity 0.87 vs LR sensitivity 0.37 and specificity 0.84
Krikov et al (2007)[107]	Graft survival	US Renal Data System (1990–1999) (n = 92 844)	Tree-based models	AUROC (0.63, 0.64, 0.71, 0.82, 0.90) for 1-, 3-, 5-, 7-, and 10-y GS
Brier et al (2003)[24]	Delayed graft function	Patient records, USA (n = 304)	Artificial Neural Network	ANN was 63.5% sensitive and 64.8% specific. LR was specificity 90.7%
Goldfarb-Rumyantzev et al (2003)[108]	Graft survival	UNOS 1990–1998 (n = 37 407)	Tree-based model	Correlation coeff. to observed data $r = 0.998$ for LR and $r = 0.984$ for TBM for 3-y GS
Petrovsky et al (2002)[23]	Graft rejection	ANZDATA Registry Database (n = 1542)	Artificial Neural Network	Correctly predicted 59% of rejection outcomes

Continued next page

TABLE 2. (Continued)

Author (year)	Prediction	Data	ML model used	Result
Liver				
Molinari et al (2019)[19]	Mortality post-liver Tx (90 d)	UNOS Registry. 2002–2013 (n = 30 458). Preoperative features	Feature extraction: ANN, tree-based models analysis	AUROC: 0.61—all patients, 0.952—patients with ≥10% predicted mortality
Wadhwani et al (2019)	Outcome post pediatric liver Tx (3 y)	SPLIT registry (n = 887). Perioperative and post-op year 1 factor considered	Tree-based model	Accuracy of 0.71, PPV = 0.83, and NPV = 0.70. This exceeds that of the naive prediction classifier
Lau et al (2017)[20]	30-d graft failure post-liver TxK	Transplant data from Australian Hospital, 2010–2013 (n = 180).	ANN, tree-based models	Tree based: AUROC = 0.818 ANN: AUROC = 0.835. Outperformed SOFT, MELD, and DRI
Khosravi et al (2015)[109]	Mortality post liver Tx	Patient data, Iran, 2008–2013 (n = 1168)	Artificial Neural Network	AUROC were 86.4% and 80.7% for ANN and CoxPH
Cruz-Ramirez et al (2013)[21]	3-mo mortality and organ allocation	Multicentre, Spain, 2007–2008 (n = 1003)	Artificial Neural Network	Developed a rule-based system for allocating donors to recipients
Cardiothoracic				
Miller et al (2019)[16]	Survival post-heart Tx (1 y)	UNOS Registry, 1987–2014 (n = 56 477)	Multiple models: ANN, SVM, tree-based models	ANN highest C-statistic = 0.66, regression 0.65
Medved et al (2019)[17]	Survival post heart Tx (1 y)	UNOS Registry, 1997–2011 (n = 27 860)	Comparison of the IHSTA and IMPACT models	1-y survival, AUROC: IHSTA (0.654) and (0.608) IMPACT models
Miller et al (2019)[110]	Survival post heart Tx (pediatric)	UNOS Registry, 2006–2015 (n = 3502)	Multiple models compared: ANN, tree-based models	1-y survival: AUROC RF: 0.72, ANN: 0.65, CART: 0.67. Sensitivity poor ranged (7%–44%)
Yoon et al (2018)[8]	Survival post-heart Tx and waiting list mortality	UNOS Registry, 1985–2015 (n = 59 820: heart transplant recipients) and (n = 35 455: listed but did not undergo heart Tx)	Tree-based model	Performed best in predicting 3-mo survival: AUC: 0.66, C-statistic 0.57 Outperformed DRI, IMPACT, RSS. Just outperformed LR
Dag et al (2016)[111]	Survival post heart Tx (9 y)	UNOS, 1987–2012 (n = 13 720)	Bayesian Belief Network	BBN method provides similar predictive performance to the best approaches in the literature
Nilsson et al (2015)[7]	Survival post-heart Tx (1 y)	International Society for Heart and Lung Transplantation registry 1994–2010 (n = 56 625). External validation NTTD (n = 1285)	Artificial Neural Network and tree-based models. Compared against to existing scoring models	IHSTA outperformed other models AUROC: IHSTA (0.65), RSS (0.61), IMPACT (0.61), and DRI (0.56) Allocation simulation proposed a 22% increase in suitable donors
Osteokine et al (2009)[18]	Outcomes following heart-lung Tx	UNOS Registry (1987–2009) (n = 16 604)	Multiple models used for feature extraction: ANN, tree-based models	ML approaches identified novel features of importance that improved the performance of CoxPH models
Delen et al (2010)[112]	Survival following lung Tx	UNOS Registry (1987–2010) (n = 106 394). Includes pre- and post-transplant features	Multiple models used for feature extraction: ANN, tree-based models, SVM	ML identified novel features. SVM mined features performed optimally with an $R^2 = 0.879$

ANN, artificial neural network; ANZDATA, Australia and New Zealand Dialysis and Transplant Registry; AUROC, area under the receiver operating curve; BBN, Bayesian Belief Network; CoxPH, Cox proportional hazards models; DRI, donor risk index; GS, graft survival; IMPACT, index for mortality prediction after cardiac transplantation; LR, logistic regression; MELD, Model for End-Stage Liver Disease; ML, machine learning; NTTD, nordic thoracic transplantation database; RSS, risk stratification score; SPLIT, studies of pediatric liver transplantation; SVM, support vector machine; TBM, tree-based models; Tx, transplant; UNOS, United Network for Organ Sharing.

variables, limiting their utility in predicting future risk following transplantation. ML models have the advantage in this regard, in that they can better handle sequences of data and missingness of data, and can be trained to extract variables (features) from complex datasets (eg, electronic health records).[1] Esteban et al[30] used a recurrent neural network architecture to predict the likelihood of renal allograft rejection, graft loss, and patient death at the point of assessment in clinic for the future 6 and 12 months. As well as considering baseline (static) features, they considered variables such as laboratory results and medication regimen as continuous variables following transplantation. The model was able to harness data from 193 111 clinic visits made by the 2061 patients included and achieved an AUROC of 0.833 and area under the precision-recall curve of 0.345 for the overall outcome prediction. Important limitations to note include the single-center nature of the study, which will affect the generalisability of the model, as the data points generated will represent elements of local practice and human decisions embedded within the quantitative data. Validation with an external cohort is required. Nonetheless, this approach could be considered to tailor treatment regimens, inform patients of their postoperative risk, and potentially guide weaning of immunosuppression.

Immunosuppression Regimen

Accurate immunosuppression dosing can be challenging for clinicians, as multiple drug-drug interactions and narrow therapeutic windows can confound predictable dose responses to therapy. This is an issue not only for optimal graft survival but also to minimize adverse events due to toxicity and minimize patients' length of stay in busy transplant units. In an effort to address this issue following transplantation, a number of groups have used ML modeling to allow more accurate tacrolimus dosing post-renal transplant when compared to a clinician's dosing decisions.[31-34] Sample sizes ranged from 80 to 1045 and each study took place in a single-center, in sites including China, India, and Norway. Importantly, none were validated on geographically distant cohorts. Furthermore, these studies focused on acute tacrolimus dosing, and it remains unknown whether such approaches would be helpful in longer terms of dosing decisions. A Bayesian cost-benefit model showed that, after a year, there was a clear benefit from calcineurin inhibitor-free plus basiliximab induction therapies, with a slight benefit from calcineurin inhibitor-sparing protocols.[35] Finally, discharge post successful liver transplant can be accelerated by optimized drug dosing using a patient-level approach: "the parabolic personalized dosing platform" is a mathematical surface represented by a second-order algebraic equation.[36]

In future, we may see AI increasingly utilized for drug discovery in transplantation. Recently an ANN has discovered a highly effective novel antibiotic termed "halicin," repurposed from another drug.[37] Trained on molecules known to be effective against *Escherichia coli*, the ANN was then able to screen other drug compounds and predict potentially effective molecules based on patterns unseen by human experts. A similar approach could be utilized for the discovery of novel immunosuppressants.

Machine Learning for Diagnostic Image Analysis

ML and DL have huge clinical potential in the analysis of histopathological and radiological images for computer-aided diagnostics. This could be particularly helpful given the out-of-hours nature of transplantation and could theoretically improve standardization of assessments across the world using cloud-based diagnostic systems and increase the size of available data for future research.

Image analysis is essentially pattern recognition and, as such, ML models can be trained to do this with tasks ranging from automated classification of pixels for the detection and segmentation (eg, identification of organs) to assisting in complex classification (eg, grades of rejection). DL models predominantly using CNNs are the most popular contemporary ML approach for image analysis, as images may be assessed without significant pre-processing. Many studies will use a transfer learning approach, whereby algorithms trained on the classification of images will be transferred and re-trained onto a new classification problem (eg, organ size on CT).[38] This is important, as clinical datasets will typically be small (eg, hundred/thousands) as compared to the pre-trained models (millions of natural images).[39]

Machine Learning in Digital Pathology

ML is now feasible in pathology with the increasing adoption of digital pathology scanners, able to capture whole slide images at microscopic resolution in just minutes. ML-assisted analysis of pathological images has the potential to increase efficiency of workflow, provide a second opinion to a pathologist and potentially improve standardization of diagnoses across regions.

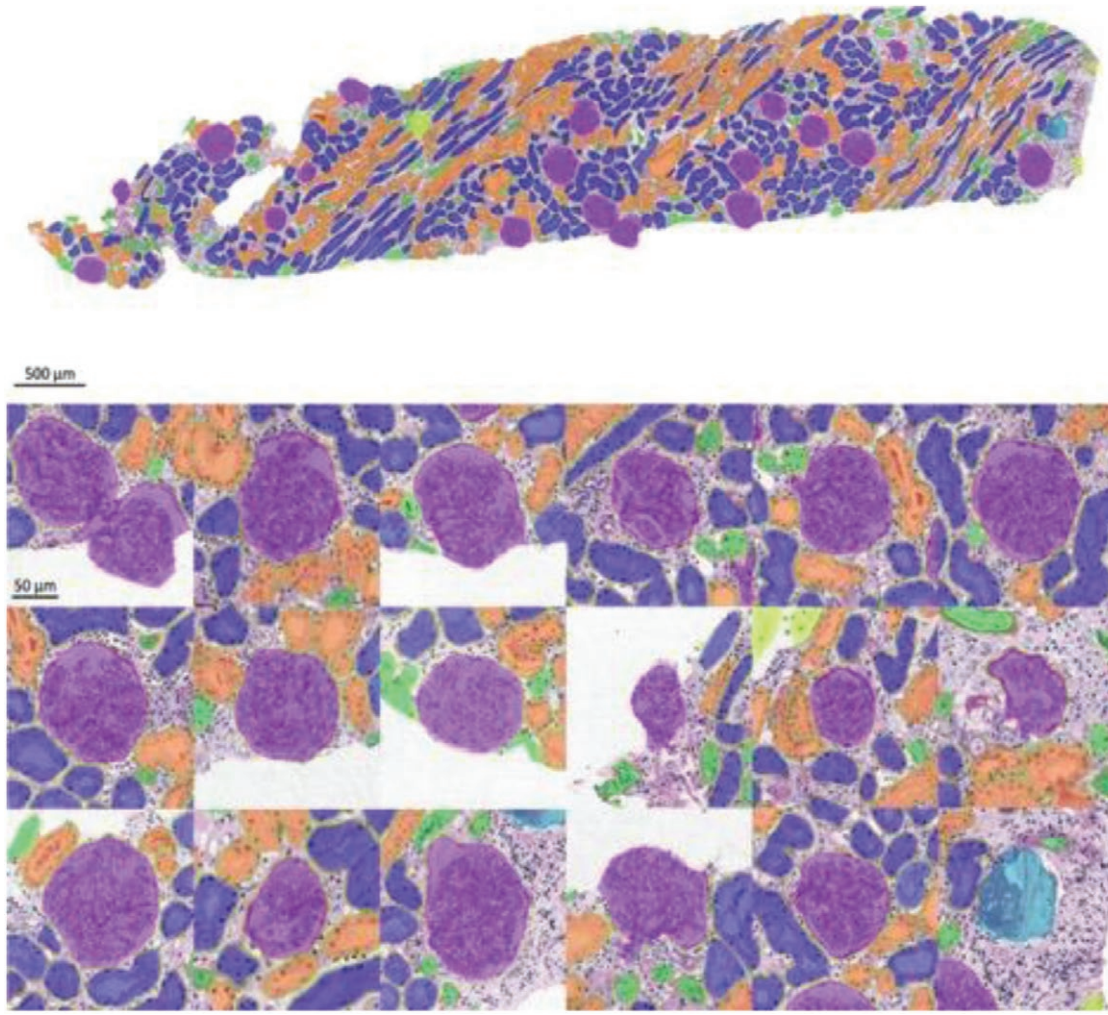
Within transplantation, renal transplant biopsies have been most extensively studied using ML models, with some success. Quantification of sclerotic glomeruli is a labor-intensive task for a pathologist and may be required rapidly and out of hours for pre-implantation biopsies. Marsh et al[40] employed a CNN for this very task and achieved a precision score of 0.8128 and 0.6070 for the detection of non-sclerosed and sclerosed glomeruli respectively, performing almost to the level of a certified pathologist. More complex structural classification to accurately segment glomeruli, interstitium and tubules of renal transplant biopsies slides and nephrectomy specimens stained has been achieved by Hermsen et al[41] (Figure 3). Having trained the model in accurate structural segmentation, they assessed the ability of the CNN to perform routine examination of 82 renal transplant biopsies using the Banff classification.[41] The model performed best in the detection of sclerotic glomeruli but struggled to identify tubular atrophy. In part, they felt that this limitation was due to the inconsistencies in the agreed diagnosis/labeling of atrophied tubules between the 3 pathologists,[38] resulting in knock-on effects on the model's performance. F3

Machine Learning in Radiological Imaging

High volume of image data for training, a well-developed IT infrastructure, and perhaps more clearly defined outcomes than pathology have positioned radiology as one of the earliest adopters of ML in healthcare. There are now FDA-approved ML models being evaluated across a range of settings such as segmentation of liver CT and mammogram analysis, with perceived benefits of increased efficiency, improved workflow, reduced intra-observer variability, and in some cases, improved performance as compared to radiologists cited.[2,39,42]

Although few studies specifically study imaging in transplantation, there are a large number of ML studies of radiological images with translational benefit. Accurate estimation of liver volume is important in the assessment for live donor liver transplantation. Normally, this task is manually performed by delineating the liver segments on each CT slice; however, CNNs have been performed to automate this labor-intensive process with high accuracy.[43] In autosomal dominant polycystic kidney disease, total kidney volume is a biomarker for disease progression.[44] ML models have been applied with success to automatically calculate the kidney and liver volume of patients with autosomal dominant polycystic kidney disease from abdominal imaging such as CT and MRI (Figure 4).[45-47] This has the potential to increase the reproducibility and efficiency of repeated measurements which may be required when assessing response to treatment. F4

Before liver transplantation, a rapid and accurate quantification of liver graft steatosis would guide the implanting team as to risks of a given donor organ. At present, the gold standard of steatosis assessment is histopathological, however, this is not routinely available before implantation. Neural networks have been developed for the automated detection of liver steatosis and fibrosis on ultrasound and this may eventually allow non-radiology



AQ11 **FIGURE 3.** Segmentation of structures on a transplant renal biopsy is achieved by a deep learning model. This convolutional neural network developed by Hermsen et al[41] was trained in the detection of glomeruli, interstitium, and tubules performed on whole slide images. The correctly identified glomeruli are depicted in the high magnification panel below. The model was unable to differentiate between the two adjacent glomeruli in the top left image. The detected sclerotic glomerulus is shown in the bottom right image in blue. This image has been adapted with permission from the publisher.

specialists to generate these measurements with reliability, for example, at the time of organ retrieval.[48,49]

Following lung transplantation, ML models have been used to predict the development of bronchiolitis obliterans syndrome. Quantitative computed tomography metrics, combined with functional respiratory imaging, was analyzed using a SVM to predict future changes in lung capacity.[50]

As well as performing image analysis, ML models can learn from the fusion of heterogenous datasets including, for example, imaging data, biomarkers, and clinical variables. For the detection of AR following renal transplantation, the combination of diffusion-weighted MRI using a CNN, and a SVM which has been combined with measures of serum creatinine clearance for the detection of AR achieved an high accuracy of 92.9%.[51]

Overall, whilst increasingly accurate segmentation of structures on pathological and radiological imaging relevant to transplantation can be achieved, the classification and diagnosis of disease lag behind. However, advances are being made that could make computer-assisted diagnostics a reality. In a landmark paper by Esteva et al,[52] a CNN was able to perform at the level of a dermatologist in the detection of skin cancers. In future, this ML model employed in a smartphone device could be part of the assessment and post-operative monitoring of our transplant patients who are at higher risk for skin malignancy; guiding the team as to the urgency of dermatology referral required.

Machine Learning in Transplantation Research

Modern biomedical research involves multiple high-throughput systems and can result in the generation of a vast quantity of complex data. From helping cluster single-cell transcriptomic data, to modeling drug-receptor interactions or quantifying high volumes of staining through microscopy, ML is now found in almost every corner of the modern laboratory. While beyond the scope of this review, it is worth noting some of the broad themes in this domain have obvious translational aims and impacts on patients and their teams. AQ3

Machine Learning in Multiomics

High-throughput technologies such as genomics, transcriptomics, and proteomics are driving biomedical discoveries in modern medical research and have required

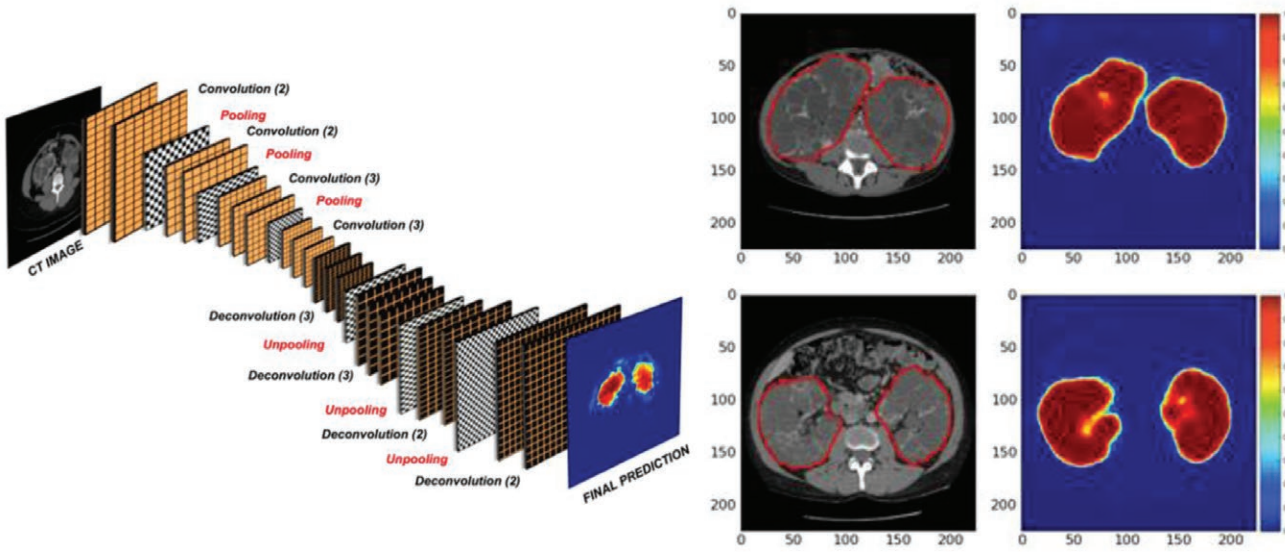


FIGURE 4. Application of a Convolutional Neural Network (CNN) in autosomal dominant polycystic kidney disease (ADPKD). A, This schematic is used as an example of the architecture of a CNN. In this study by Sharma et al, 10 fully convolutional layers were used to progressively modify and compress the spatial size of the pixels from the original image. Pixelwise segmentation was then employed to deconvolute the data and generate the final prediction heatmaps. B, Predicted kidney segmentation contour (red color) based on the adjacent CT scans from 2 patients with ADPKD. Images used with permission from the journal.[46]

multiple advances in computational biology and bioinformatics. Almost all work using these technologies will rely on some degree of ML for the analysis of the data generated, from the moment it is downloaded from the sequencer, through clustering methods, statistical comparisons, and more advanced analyses.

As an example applied to transplantation research, single nucleotide polymorphisms associated with long-term clinical outcome in renal transplant patients have been identified using genome-wide association studies. The authors employed multiple ML approaches in the Plink package before using multivariable Cox regression analyses to explore their single nucleotide polymorphism associations.[53,54]

AQ4 In transcriptomics, ML can be applied to bulk RNASEQ, and has allowed deconvolution of the lymphocyte compartment of the peripheral whole blood transcriptome in the context of acute kidney allograft rejection in an effort to predict outcomes.[55] In the realm of single-cell transcriptomics, ML approaches are fundamental to the workflow. For example, in exploring the Human Kidney Allograft inflammatory response at a single cell level, authors used ML at every stage of the analysis, from aligning the raw reads to a transcriptome, through to clustering (k-mean shared nearest neighbor approach), dimensionality reduction (T-distributed stochastic neighbor embedding) and trajectory analysis (reversed graph embedding and DDRTree).[56]

Finally, linear discriminant analyses and SVMs have been used successfully to classify renal biopsy samples as having T-cell mediated rejection. Based on the top differentially expressed genes in paraffin-embedded sections, this represents a novel advancement on traditional methods that relied upon pathological analysis of the biopsy itself.[25] Ongoing work in transplantation to create a "molecular microscope"—that is, to identify molecular signatures of specific histological diagnosis such as graft rejection in both biopsy samples and peripheral blood samples—rely heavily on a number of ML methods to analyze the high volumes of microarray and RNA-sequencing required.[56-63]

In future, ML may be able to incorporate omics data with electronic health record data, with the promise that this will provide a path to a precision approach in our transplant patients. There are steps towards this, for example, with the Electronic Medical Records and Genomics Network consortium is now established which aims to leverage electronic health records as a tool for genomic research.[64]

In Silico Trials

In silico trials are computer simulations that allow modeling of biological systems to support clinical trial design and interpretation. Modeling simulated diseases in a thousand clinical trials allowed researchers to compare multiple trial methodologies using an in silico approach, including important characteristics such as power, precision, duration, and number of patients needed to recruit.[65] Such an approach would have obvious benefits in optimizing trials in solid organ transplantation, where recruiting patients can be challenging due to small populations to draw from, thus requiring trials to run for longer and at greater expense to achieve endpoints and required power. Trial safety could be enhanced in this way: routine and novel markers can be modeled in silico to allow more sensitive and specific detection of important safety endpoints in trials, thus increasing the detection of important adverse events and reducing their impact with earlier detection, as has been the case in drug-induced liver injury.[66,67]

CHALLENGES IN CLINICAL TRANSLATION OF ML MODELS

Despite the growth of ML tools in transplantation, few have made the transition into real-world clinical application.[68] A clinically useful ML model needs to be accurate, transparent, generalizable, regulated, and must crucially

result in improved clinical outcomes. There are both limitations inherent to ML models and human factors that must be addressed.

Limitations of ML Models: Bias, Brittleness, and the Black Box

Bias

To achieve accurate predictions, training on large, high-quality datasets is required. ML models will use any aspect of the data to achieve the best possible predictions and this can result in unintended biases. When Amazon deployed an AI model trained to screen resumés of potential job applicants based on the features of previous successful applicants, who were predominantly male, the model had learned to discriminate against female applicants. Another deep model appeared able to accurately discriminate the difference between pictures of dogs and wolves, only to discover the way it did this was by noting that wolves usually were standing on snow rather than grass in the photographs.[69]

Representative biases can occur in genetic databases, clinical databases, and social databases, and are potential pitfalls regardless of the ML approach used.[70-74] In transplantation, we must be mindful that ethnic minority groups are appropriately represented, as imbalances in the donor and recipient pools and organ outcomes are known to exist.[75] Models may be biased towards the detection of more common conditions, potentially to the detriment of rarer conditions, particularly in the training of ANNs for image analysis that require huge datasets. Additionally, most ML models are trained on retrospective/historical datasets that may not reflect current practice. Changes in clinical practice—for example, organ allocation, left ventricular assist device introduction, and immunosuppression—may result in "data shift," which could result in unintended bias or a reduction in model performance when applied prospectively.

Brittleness

Generalisability remains a key barrier to the clinical translation of most ML models and results in reduced performance of the ML model when applied to external data. As models become increasingly complex, they also become "brittle" to variations in input data that are inherent between centers, for example, variance in coding definitions, laboratory ranges, image acquisition, and diagnostic classification. In the context of image analysis, the performance of many models in study conditions can outperform diagnostician performance, but when generalized to real-world data the performance of these models is likely to drop or, occasionally, to fail spectacularly.[76,77] To avoid errors, a CNN must be trained in the accurate analysis of "sub-optimal" images. In pathology, this would be the ability to account for artefactual differences including tissue orientation, staining quality, staining type and specimen folding, photographic quality, and microscope focus. Concerningly, few of the studies reviewed here performed any validation on external datasets. This is not unusual; in a systematic review of AI for diagnostic medical imaging, just 6% performed external validation.[68,78]

However, this is not insurmountable. Testing and recalibrating a model using previously unseen data from external sources will help counter overfitting to a particular dataset and may mitigate certain biases by increasing the diversity of the patient demographic studied. Identification of "input outliers" before analysis is another strategy to reduce potential errors. International collaboration to provide the necessarily large and diverse datasets needed to mitigate the risks of bias and brittleness will be key for the future of ML in transplantation.

The Black Box

In many circumstances, there remains a trade-off between the performance capabilities of an ML model and the model's transparency, or lack thereof. Deep models often have high predictive capabilities with poor transparency as compared to models that are inherently explainable (eg, random forests, stochastic gradient boosting) but which may have poorer predictive capabilities. Intense research is focused on providing explicability of such models—for example, working backwards to determine the relative weighting given to variables within a model—but it is often extremely difficult to rationalize a decision that was made in the "black box." Given the unintended biases and brittleness that can occur, such transparency would be essential before a ML model could be implemented for the high-stakes' decision-making common in clinical transplantation.[79]

Assuming the correct diagnosis is achieved by the ML model, clinicians must have confidence in its output for it to have value in decision making. Providing a window into the "black box" is one way to garner trust. To illustrate, in the context of palliative care, until the intermediate steps of a neural network were visible and interpretable to clinicians, they were reticent to be influenced by the predictions and recommendations produced by the model.[80] Whilst in digital pathology, rather than simply providing a binary disease output (eg, cancer, no cancer), ML can now furnish textual descriptions (eg, "nucleic pleomorphisms") for the images of perceived abnormal areas that contributed to its diagnoses in bladder cancer cases.[81] Such steps are likely to improve the relationship between man and machine, and to achieve this we must work closely with data scientists to achieve models that may be adopted clinically.

Human Barriers to Clinical Adoption

If an ML model is to be applied to our patients, it must be ethically sound, clinically effective and quality assured by regulation.

Given the propagated biases and brittleness that can occur, there are concerns as to whether complex ML models have the necessary transparency to allow rigorous evaluation, regulation, and to be explicable to multiple stakeholders.[82] Such transparency is essential to prove that models are fair, particularly if they are ever to be applied to high-stakes' decisions such as organ allocation. In lower stakes' decisions, that is, the clustering of molecular data or in CT analysis in which a second independent assessment will be performed, the need for transparency may be lower.

Our understanding of the ethical complexities of adopting AI models into healthcare currently lags behind the capabilities of these models. There are pertinent concerns around data ownership, security, consent, and confidentiality in data used to train ML models.[83] Furthermore, as roles of the machine and the physician become intertwined,

we must ensure that patient preference is taken into consideration. Can a physician or patient be sure that the AI model was trained on a patient population in which they would be adequately represented? Are patients comfortable with the potential opacity of AI models? What happens when things go wrong? Physicians must have sufficient understanding of AI to be able to engage in this ethical dialogue. The development of ethical frameworks is underway to help address these issues.[84-86]

In assessing the clinical effectiveness of AI models, the first step is to ensure that studies published are of high quality. Education in medical school and familiarization of reviewers in the critical appraisal of ML studies is a key step and we would recommend the following articles for further reading.[80,87,88] The recent development of the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis in ML framework should help by standardizing the reporting of prognostic ML models in future.[89] In addition, diagnostic predictive models should be tested in randomized controlled trials before clinical implementation, as high model accuracy does not necessarily equate to clinical benefit.[68,90]

To safeguard clinical efficacy and safety, regulation is required. Models influencing disease diagnosis, prevention, or treatment are categorized as "software as a medical devices."[91] As medical devices, they require the FDA approval or CE marking in the US and Europe, respectively. The degree of regulation required is graded according to a number of factors, including the potential clinical impact of the model output.

One particular challenge specific to the ML models is that, as the data naturally changes over time or is applied to different populations, recalibration is required to ensure optimum performance. This would be key in transplantation, where clinical practice rapidly evolves and significant international differences in practice exist. However, these recalibrated models will be subtly different from the model initially validated and this poses regulatory and ethical challenges. Specific guidance is still being developed with such challenges in mind.[91,92]

CONCLUSION

Over the last 10 years, we have witnessed astonishing advances in ML models, which are now driving these approaches to the brink of clinical translation in transplantation. Already, ML models are tackling complex predictions with at least comparable accuracy to traditional models, assisting in everything from organ allocation to immunosuppression outcomes. The ability to analyze more heterogenous, longitudinal, and/or fused datasets (eg, genomics plus, electronic health records, and imaging) is likely to be where ML provides clear advantages over traditional approaches. In image analysis, automated segmentation for generation of metrics for tasks that are otherwise labor-intensive appears promising.

There are key technical and ethical challenges that must be addressed before ML enters the clinical realm, but these are not insurmountable. Engagement of the transplant community with data scientists and engineers to ensure that the data provided to train these AI algorithms will enhance the accuracy of AI in clinical decision support to ultimately improve transplant outcomes.

REFERENCES

1. Tomašev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*. 2019;572:116–119.
2. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25:30–36.
3. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet.* 2019;394:861–867.
4. Zarins CK, Taylor CA, Min JK. Computed fractional flow reserve (FFTCT) derived from coronary CT angiography. *J Cardiovasc Transl Res.* 2013;6:708–714.
5. Chollet F, Allaire J. *Deep Learning With R*. Manning Publications; 2018. AQ5
6. Bertsimas D, Kung J, Trichakis N, et al. Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation. *Am J Transplant.* 2019;19:1109–1118.
7. Nilsson J, Ohlsson M, Höglund P, et al. The International Heart Transplant Survival Algorithm (IHTSA): A new model to improve organ sharing and survival. *PLoS One.* 2015;10:e0118644.
8. Yoon J, Zame WR, Banerjee A, et al. Personalized survival predictions via trees of predictors: An application to cardiac transplantation. *PLoS One*. 2018;13:e0194985.
9. Davis AE, Mehrotra S, Friedewald JJ, et al. Improving geographic equity in kidney transplantation using alternative kidney sharing and optimization modeling. *Med Decis Making.* 2015;35:797–807.
10. Magruder JT, Shah AS, Crawford TC, et al. Simulated regionalization of heart and lung transplantation in the United States. *Am J Transplant.* 2017;17:485–495.
11. Ayllón MD, Ciria R, Cruz-Ramírez M, et al. Validation of artificial neural networks as a methodology for donor-recipient matching for liver transplantation. *Liver Transpl.* 2018;24:192–203.
12. Al-Ebbini L, Oztekin A, Chen Y. FLAS: Fuzzy lung allocation system for US-based transplantations. *Eur J Oper Res*. 2016;248:1051–1065.
13. Arrieta-Bolaños E, Crivello P, Shaw BE, et al. In silico prediction of nonpermissive HLA-DPB1 mismatches in unrelated HCT by functional distance. *Blood Adv.* 2018;2:1773–1783.
14. Nursetyo AA, Syed-Abdul S, Uddin M, et al. Graft rejection prediction following kidney transplantation using machine learning techniques: A systematic review and meta-analysis. *Stud Health Technol Inform.* 2019;264:10–14.
15. Khush KK, Zaroff JG, Nguyen J, et al. National decline in donor heart utilization with regional variability: 1995-2010. *Am J Transplant*. 2015;15:642–649.
16. Miller PE, Pawar S, Vaccaro B, et al. Predictive abilities of machine learning techniques may be limited by dataset characteristics: Insights from the UNOS database. *J Card Fail.* 2019;25:479–483.
17. Medved D, Ohlsson M, Höglund P, et al. Improving prediction of heart transplantation outcome using deep learning techniques. *Sci Rep.* 2018;8:3613.
18. Oztekin A, Delen D, Kong ZJ. Predicting the graft survival for heart-lung transplantation patients: An integrated data mining methodology. *Int J Med Inform.* 2009;78:e84–e96.
19. Molinari M, Ayloo S, Tsung A, et al. Prediction of perioperative mortality of cadaveric liver transplant recipients during their evaluations. *Transplantation.* 2019;103:e297–e307.
20. Lau L, Kankanige Y, Rubinstein B, et al. Machine-learning algorithms predict graft failure after liver transplantation. *Transplantation.* 2017;101:e125–e132.
21. Cruz-Ramírez M, Hervás-Martínez C, Fernández JC, et al. Predicting patient survival after liver transplantation using evolutionary multi-objective artificial neural networks. *Artif Intell Med*. 2013;58:37–49.
22. Lee HC, Yoon S, Yang SM, et al. Prediction of acute kidney injury after liver transplantation: Machine learning approaches vs. logistic regression model. *J Clin Med*. 2018;7:428.
23. Petrovsky N, Tam SK, Brusic V, et al. Use of artificial neural networks in improving renal transplantation outcomes. *Graft*. 2002;5:6.
24. Brier ME, Ray PC, Klein JB. Prediction of delayed renal allograft function using an artificial neural network. *Nephrol Dial Transplant.* 2003;18:2655–2659.
25. Santori G, Fontana I, Valente U. Application of an artificial neural network model to predict delayed decrease of serum creatinine in pediatric patients after kidney transplantation. *Transplant Proc*. 2007;39:1813–1819.
26. Akl A, Ismail AM, Ghoneim M. Prediction of graft survival of living-donor kidney transplantation: Nomograms or artificial neural networks? *Transplantation.* 2008;86:1401–1406.

27. Hummel AD, Maciel RF, Rodrigues RG, et al. Application of artificial neural networks in renal transplantation: Classification of nephrotoxicity and acute cellular rejection episodes. *Transplant Proc.* 2010;42:471–472.
28. Loupy A, Aubert O, Orandi BJ, et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: International derivation and validation study. *BMJ.* 2019;366:l4923.
29. Pallardó Mateu LM, Sancho Calabuig A, Capdevila Plaza L, et al. Acute rejection and late renal transplant failure: Risk factors and prognosis. *Nephrol Dial Transplant*. 2004;19(Suppl 3):iii38–iii42.
30. Esteban C, Staeck O, Yang Y, Tresp V. Predicting clinical events by combining static and dynamic information using recurrent neural networks. 2016.
31. Størset E, Åsberg A, Skauby M, et al. Improved tacrolimus target concentration achievement using computerized dosing in renal transplant recipients—A prospective, randomized study. *Transplantation.* 2015;99:2158–2166.
32. Tang J, Liu R, Zhang YL, et al. Application of machine-learning models to predict tacrolimus stable dose in renal transplant recipients. *Sci Rep*. 2017;7:42192.
33. Niel O, Bastard P. Artificial intelligence improves estimation of tacrolimus area under the concentration over time curve in renal transplant recipients. *Transpl Int.* 2018;31:940–941.
34. Thishya K, Vattam KK, Naushad SM, et al. Artificial neural network model for predicting the bioavailability of tacrolimus in patients with renal transplantation. *PLoS One.* 2018;13:e0191921.
35. Emparan C, Wolters H, Laukötte M, et al. The cost-effectiveness of basiliximab induction in "old-to-old" kidney transplant programs: Bayesian estimation, simulation, and uncertainty analysis. *Transplant Proc.* 2005;37:2069–2071.
36. Zarrinpar A, Lee DK, Silva A, et al. Individualizing liver transplant immunosuppression using a phenotypic personalized medicine platform. *Sci Transl Med.* 2016;8:333ra49.
37. Stokes JM, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery. *Cell.* 2020;180:688–702.e13.
38. Kannan S, Morgan LA, Liang B, et al. Segmentation of glomeruli within trichrome images using deep learning. *Kidney Int Rep.* 2019;4:955–962.
39. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60–88.
40. Marsh JN, Matlock MK, Kudose S, et al. Deep learning global glomerulosclerosis in transplant kidney frozen sections. *IEEE Trans Med Imaging.* 2018;37:2718–2728.
41. Hermsen M, de Bel T, den Boer M, et al. Deep learning-based histopathologic assessment of kidney tissue. *J Am Soc Nephrol.* 2019;30:1968–1979.
42. Le EPV, Wang Y, Huang Y, et al. Artificial intelligence in breast imaging. *Clin Radiol.* 2019;74:357–366.
43. Lu F, Wu F, Hu P, et al. Automatic 3D liver location and segmentation via convolutional neural network and graph cut. *Int J Comput Assist Radiol Surg.* 2017;12:171–182.
44. Grantham JJ, Torres VE, Chapman AB, et al. Volume progression in polycystic kidney disease. *N Engl J Med*. 2006;354:2122–2130.
45. van Gastel MDA, Edwards ME, Torres VE, et al. Automatic measurement of kidney and liver volumes from MR images of patients affected by autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*. 2019;30:1514–1522.
46. Sharma K, Rupprecht C, Caroli A, et al. Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease. *Sci Rep*. 2017;7:2049.
47. Simms RJ, Doshi T, Metherall P, et al. A rapid high-performance semi-automated tool to measure total kidney volume from MRI in autosomal dominant polycystic kidney disease. *Eur Radiol*. 2019;29:4188–4197.
48. Byra M, Styczynski G, Szmigielski C, et al. Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. *Int J Comput Assist Radiol Surg.* 2018;13:1895–1903.
49. Cassinotto C, Pierredon-Foulongne MA, Belgour A, et al. Learning curve of liver stiffness measurement using a new hybrid machine composed of transient elastography interfaced with ultrasound. *Eur Radiol.* 2020;30:1088–1095.
50. Barbosa EJM Jr, Lanclus M, Vos W, et al. Machine learning algorithms utilizing quantitative CT features may predict eventual onset of bronchiolitis obliterans syndrome after lung transplantation. *Acad Radiol.* 2018;25:1201–1212.
51. Abdeltawab H, Shehata M, Shalaby A, et al. A novel CNN-based CAD system for early assessment of transplanted kidney dysfunction. *Sci Rep.* 2019;9:5948.
52. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–118.
53. Pihlstrøm HK, Mjøen G, Mucha S, et al. Single nucleotide polymorphisms and long-term clinical outcome in renal transplant patients: A validation study. *Am J Transplant*. 2017;17:528–533.
54. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–575.
55. Shannon CP, Balshaw R, Ng RT, et al. Two-stage, in silico deconvolution of the lymphocyte compartment of the peripheral whole blood transcriptome in the context of acute kidney allograft rejection. *PLoS One.* 2014;9:e95224.
56. Wu H, Malone AF, Donnelly EL, et al. Single-cell transcriptomics of a human kidney allograft biopsy specimen defines a diverse inflammatory response. *J Am Soc Nephrol.* 2018;29:2069–2080.
57. Christakoudi S, Runglall M, Mobillo P, et al. Development of a multivariable gene-expression signature targeting T-cell-mediated rejection in peripheral blood of kidney transplant recipients validated in cross-sectional and longitudinal samples. *Ebiomedicine.* 2019;41:571–583.
58. Venner JM, Famulski KS, Badr D, et al. Molecular landscape of T cell-mediated rejection in human kidney transplants: prominence of CTLA4 and PD ligands. *Am J Transplant.* 2014;14:2565–2576.
59. Halloran PF, Reeve JP, Pereira AB, et al. Antibody-mediated rejection, T cell-mediated rejection, and the injury-repair response: new insights from the Genome Canada studies of kidney transplant biopsies. *Kidney Int.* 2014;85:258–264.
60. Kurian SM, Velazquez E, Thompson R, et al. Orthogonal comparison of molecular signatures of kidney transplants with subclinical and clinical acute rejection: Equivalent performance is agnostic to both technology and platform. *Am J Transplant.* 2017;17:2103–2116.
61. Moulavi D, Hajiloo M, Sander J, et al. Combining gene expression and interaction network data to improve kidney lesion score prediction. *Int J Bioinform Res Appl.* 2012;8:54–66.
62. Liu P, Tseng G, Wang Z, et al. Diagnosis of T-cell-mediated kidney rejection in formalin-fixed, paraffin-embedded tissues using RNA-Seq-based machine learning algorithms. *Hum Pathol.* 2019;84:283–290.
63. Reeve J, Böhmig GA, Eskandary F, et al. Assessing rejection-related disease in kidney transplant biopsies based on archetypal analysis of molecular phenotypes. *JCI insight*. 2017;2:e94197.
64. Gottesman O, Kuivaniemi H, Tromp G, et al.; eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: Past, present, and future. *Genet Med.* 2013;15:761–771.
65. Bajard A, Chabaud S, Cornu C, et al.; CRESim & Epi-CRESim study groups. An in silico approach helped to identify the best experimental design, population, and outcome for future randomized clinical trials. *J Clin Epidemiol.* 2016;69:125–136.
66. Church RJ, Watkins PB. In silico modeling to optimize interpretation of liver safety biomarkers in clinical trials. *Exp Biol Med (Maywood).* 2018;243:300–307.
67. Watkins PB. The DILI-sim initiative: insights into hepatotoxicity mechanisms and biomarker interpretation. *Clin Transl Sci.* 2019;12:122–129.
68. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17:195.
69. Riberiro M, Singh S, Guestrin C. Why should I trust you?: Explaining the predictions of any classifier. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. 2016.
70. Spratt DE, Chan T, Waldron L, et al. Racial/ethnic disparities in genomic sequencing. *JAMA Oncol.* 2016;2:1070–1074.
71. Dakhoul L, Gawrieh S, Jones KR, et al. Racial disparities in liver transplantation for hepatocellular carcinoma are not explained by differences in comorbidities, liver disease severity, or tumor burden. *Hepatol Commun.* 2019;3:52–62.
72. Tjaden LA, Noordzij M, van Stralen KJ, et al.; ESPN/ERA-EDTA Registry Study Group. Racial disparities in access to and outcomes of kidney transplantation in children, adolescents, and young adults: Results from the ESPN/ERA-EDTA (European Society of Pediatric Nephrology/European Renal Association-European Dialysis and Transplant Association) Registry. *Am J Kidney Dis.* 2016;67:293–301.
73. McCarthy AM, Bristol M, Domchek SM, et al. Health care segregation, physician recommendation, and racial disparities in BRCA1/2 testing among women with breast cancer. *J Clin Oncol.* 2016;34:2610–2618.
74. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: Retrospective observational study. *BMJ.* 2018;361:k1479.
75. Rudge C, Johnson RJ, Fuggle SV, et al.; Kidney and Pancreas Advisory Group, UK Transplant NHS BT. Renal transplantation in the United Kingdom for patients from ethnic minorities. *Transplantation.* 2007;83:1169–1173.

76. Yosinski J, Clune J. *Deep Neural Networks are Easily Fooled : High Confidence Predictions for Unrecognizable Images*.
77. Heaven D. Why deep-learning AIs are so easy to fool. *Nature.* 2019;574:163–166.
78. Kim DW, Jang HY, Kim KW, et al. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers. *Korean J Radiol*. 2019;20:405–410.
79. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* 2019;1:206–215.
80. Kolachalama VB, Garg PS. Machine learning and medical education. *NPJ Digit Med.* 2018;1:54.
81. Zhang Z, Chen P, McGough M, et al. Pathologist-level interpretable whole-slide cancer diagnosis with deep learning. *Nat Mach Intell*. 2019;1:236–245.
82. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—Addressing ethical challenges. *N Engl J Med.* 2018;378:981–983.
83. Safdar NM, Banja JD, Meltzer CC. Ethical considerations in artificial intelligence. *Eur J Radiol.* 2020;122:108768.
84. Sullivan HR, Schweikart SJ. Are current tort liability doctrines adequate for addressing injury caused by AI? *AMA J Ethics*. 2019;21:160–166.
85. Celie KB, Prager K, Chaet D, et al. AMA Journal of Ethics. *Clin Ethics*. 2016;18:473–563.
86. Crigger E, Khoury C. Number 2: E188-191 MEDICINE AND SOCIETY Making policy on augmented intelligence in health care. *AMA J Ethics*. 2019;21:188–191.
87. Doshi-Velez F, Perlis RH. Evaluating machine learning articles. *JAMA.* 2019;322:1777–1779.
88. Liu Y, Chen PC, Krause J, et al. How to read articles that use machine learning: Users' guides to the medical literature. *JAMA.* 2019;322:1806–1816.
89. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet.* 2019;393:1577–1579.
90. Brocklehurst P, Field D, Greene K, et al. Computerised interpretation of fetal heart rate during labour (INFANT): A randomised controlled trial. *Lancet*. 2017;389:1719–1729.
91. Minssen T, Gerke S, Aboy M, et al. Regulatory responses to medical machine learning. *J Law Biosci*. 2020;7:1–18.
92. Food and Drug Administration. *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)*. FDA; 2019.
93. Yoo KD, Noh J, Lee H, et al. A machine learning approach using survival statistics to predict graft survival in kidney transplant recipients: A multicenter cohort study. *Sci Rep.* 2017;7:8904.
94. Mark E, Goldsman D, Gurbaxani B, et al. Using machine learning and an ensemble of methods to predict kidney transplant survival. *PLoS One.* 2019;14:e0209068.
95. Shaikhina T, Lowe D, Daga S, et al. Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. *Biomed Signal Process Control*. 2019;52:456–462.
96. Topuz K, Zengul FD, Dag A, et al. Predicting graft survival among kidney transplant recipients: A Bayesian decision support model. *Decis Support Syst*. 2018;106:97–109.
97. Tapak L, Hamidi O, Amini P, et al. Prediction of kidney graft rejection using artificial neural network. *Healthc Inform Res*. 2017;23:277–284.
98. Nematollahi M, Akbari R, Nikeghbalian S, et al. Classification models to predict survival of kidney transplant recipients using two intelligent techniques of data mining and logistic regression. *Int J Organ Transplant Med.* 2017;8:119–122.
99. Shahmoradi L, Langarizadeh M, Pourmand G, et al. Comparing three data mining methods to predict kidney transplant survival. *Acta Inform Med.* 2016;24:322–327.
100. Fouad M, Ellatif MMA, Hagag M, Akl A. Prediction of long term living donor kidney graft outcome: Comparison between rule based decision tree and linear regression. *Int J Adv Comp Res*. 2015;3:185–192.
101. Decruyenaere A, Decruyenaere P, Peeters P, et al. Prediction of delayed graft function after kidney transplantation: comparison between logistic regression and machine learning methods. *BMC Med Inform Decis Mak.* 2015;15:83.
102. Brown TS, Elster EA, Stevens K, et al. Bayesian modeling of pretransplant variables accurately predicts kidney graft survival. *Am J Nephrol.* 2012;36:561–569.
103. Lasserre J, Arnold S, Vingron M, et al. Predicting the outcome of renal transplantation. *J Am Med Inform Assoc.* 2012;19:255–262.
104. Li J, Serpen G, Selman S, Franchetti M, Riesen M, Schneider C. Bayes net classifiers for prediction of renal graft status and survival period. *World Acad Sci Eng Technol*. 2010;39. AQ8
105. Lofaro D, Maestripieri S, Greco R, et al. Prediction of chronic allograft nephropathy using classification trees. *Transplant Proc.* 2010;42:1130–1133.
106. Greco R, Papalia T, Lofaro D, et al. Decisional trees in renal transplant follow-up. *Transplant Proc.* 2010;42:1134–1136.
107. Krikov S, Khan A, Baird BC, et al. Predicting kidney transplant survival using tree-based modeling. *ASAIO J.* 2007;53:592–600.
108. Goldfarb-Rumyantzev AS, Scandling JD, Pappas L, et al. Prediction of 3-yr cadaveric graft survival based on pre-transplant variables in a large national dataset. *Clin Transplant.* 2003;17:485–497.
109. Khosravi B, Pourahmad S, Bahreini A, et al. Five years survival of patients after liver transplantation and its effective factors by neural network and Cox proportional hazard regression models. *Hepat Mon.* 2015;15:e25164.
110. Miller R, Tumin D, Cooper J, et al. Prediction of mortality following pediatric heart transplant using machine learning algorithms. *Pediatr Transplant.* 2019;23:e13360.
111. Dag A, Topuz K, Oztekin A, et al. A probabilistic data-driven framework for scoring the preoperative recipient-donor heart transplant survival. *Decis Support Syst*. 2016;86:1–12.
112. Delen D, Oztekin A, Kong ZJ. A machine learning-based approach to prognostic analysis of thoracic transplantations. *Artif Intell Med.* 2010;49:33–42.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1—For indexing purposes, please confirm that author names have been correctly identified as given names (blue) and surnames (red). Color in the byline will not appear on the final published version.

AQ2—If you have color in your proof, please indicate whether you approve the color charge by returning the color agreement with your corrections. The color agreement can be found at http://links.lww.com/TP/B705. The rate is $100 per 1 printed color figure. You may also chose to have online only color at no extra charge. If you have any questions, please contact Anna.butrim@wolterskluwer.com.

AQ3—Please check the sentence "From helping cluster single-cell transcriptomic …for clarity.

AQ4—Please define abbreviation "RNASEQ" in the article.

AQ5—Please provide publisher's location for the references 5 and 92.

AQ6—Please provide complete details for the references 30 and 76.

AQ7—Please provide conference location for reference 69.

AQ8—Please provide page range for reference 104.

AQ9—Please provide Department/Unit (if any) for all affiliations.

AQ10—Please confirm the disclosure statement.

AQ11—Please check figures 3 and 4 are low quality.